Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries

Eskil Söderlind[1]*, Leif Strandberg[1], Pernilla Jirholt[2], Norihiro Kobayashi[2,3], Vessela Alexeiva[1], Anna-Maria Åberg[1], Anna Nilsson[1], Bo Jansson[1], Mats Ohlin[2], Christer Wingren[2], Lena Danielsson[1], Roland Carlsson[1], and Carl A.K. Borrebaeck[2]

*[1]BioInvent Therapeutic AB, SE-223 70 Lund, Sweden. [2]Department of Immunotechnology, Lund University, SE-220 07 Lund, Sweden. [3]Present address: Laboratory of Bioanalytical Chemistry, Graduate School of Pharmaceutical Sciences, Tohoku University, Aobayama, Aoba-ku, Sendai 980-8578, Japan. *Corresponding author (es@bioinvent.com).*

Received 23 February 2000; accepted 8 May 2000

We constructed a single-chain Fv antibody library that permits human complementarity-determining region (CDR) gene fragments of any germline to be incorporated combinatorially into the appropriate positions of the variable-region frameworks VH-DP47 and VL-DPL3. A library of 2×10^9 independent transformants was screened against haptens, peptides, carbohydrates, and proteins, and the selected antibody fragments exhibited dissociation constants in the subnanomolar range. The antibody genes in this library were built on a single master framework into which diverse CDRs were allowed to recombine. These CDRs were sampled from in vivo-processed gene sequences, thus potentially optimizing the levels of correctly folded and functional molecules, and resulting in a molecule exhibiting a lower computed immunogenicity compared to naive immunoglobulins. Using the modularized assembly process to incorporate foreign sequences into an immunoglobulin scaffold, it is possible to vary as many as six CDRs at the same time, creating genetic and functional variation in antibody molecules.

Keywords: Antibody engineering, recombination, diversity, in vitro evolution

Highly diverse molecular libraries have recently become a vital tool in the search for molecules with novel properties. The antibody molecule in particular has been the focus of extensive development of different combinatorial strategies in order to obtain antibodies with novel specificities. These strategies include the reassembly of naturally occurring genes encoding the heavy- and light-chain domains from either immune[1,2] or nonimmune B-cell sources[3], or introduction of synthetic variability[4–10] into the individual complementarity-determining regions (CDRs), i.e. the hypervariable parts of the variable domains that interact with an antigen.

Previously we demonstrated an approach to introduce genetic diversity in antibody libraries, where diverse natural CDRs are combined into an antibody framework[11], using only CDR-encoding gene fragments originating from sequences formed and "proofread" in vivo. Such proofread peptide sequences that have evolved and developed in vivo in a specific biological context may have superior functional qualities compared with those produced in vitro. A high degree of functional variation is achieved by means of simultaneous and random combination of six biologically derived CDRs, which can extend the genetic diversity beyond what is naturally created in the immune system. In addition, the single master framework into which these CDRs are introduced was selected to be compatible with the bacterial expression system employed. Finally, the modular design of the approach allows for iterative CDR shuffling to improve specificities and affinities, and the contribution of individual CDRs can be analyzed by systematic regional changes.

In order to demonstrate the functionality of this concept, we have constructed a large human antibody single-chain fragment (scFv) library using these principles. This library has been successfully used for the isolation of antibody fragments specific for a number of different ligands, including peptides, proteins (of human and nonhuman origin), carbohydrates, and haptens. These antibodies exhibit dissociation constants in the subnanomolar range. The data clearly demonstrate that a single-antibody framework can harness diverse CDR sequences previously not found in combination, including CDR loops with foreign canonical structures, and still retain the functionality of the antibody structure.

Results and discussion

Library design using proofread gene segments. We have constructed an scFv antibody fragment gene library of 2×10^9 members, designated n-CoDeR. The genes were built using the VH DP47 and the VL DPL3 immunoglobulin genes as single master frameworks[12], since they are (1) well expressed/displayed in bacterial/phage systems[6] and (2) present in vivo in most human individuals[13–16]. Sequences encoding in vivo-formed CDRs from rearranged immunoglobulin genes of different germline origin were combined into the master framework. This was achieved by amplification of CDRs with primers specific for the DP47 and DPL3 frameworks[11] and overlap extension of single-stranded DNA (ssDNA) to create antibody genes encoding variable light and heavy chains[9,11,12]. The principle is outlined (Fig. 1).

An extremely large genetic variation can be created by random combination of six in vivo-formed CDR sequences, although an exact calculation of the variation cannot be performed, since human germline genes acquire additional genetic variation in vivo from somatic mutation processes and V(D)J recombination. The assembled scFv were cloned into the pFAB5c-His display vector[17] and

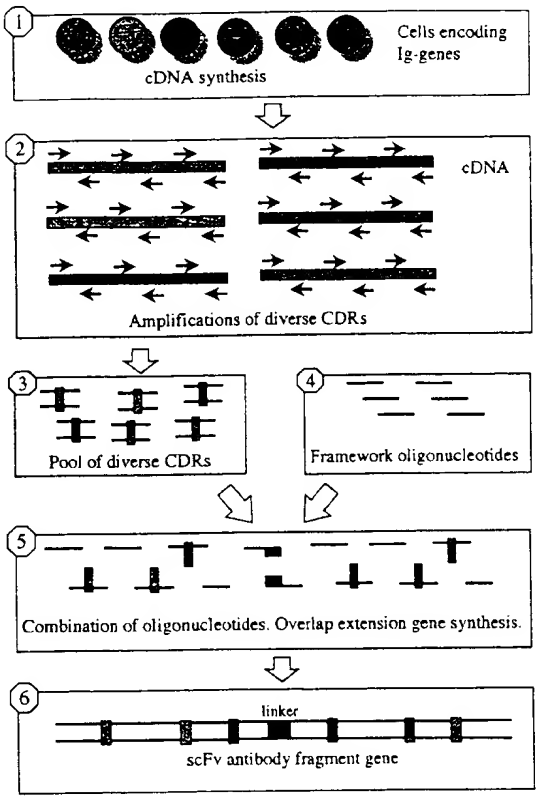

Figure 1. Schematic diagram for the construction of n-CoDeR, using in vivo-proofread gene segments. cDNA is produced (step 1) from IgM -G and -A producing cells, and this cDNA (step 2) is used as template for the specific amplification of each CDR individually (step 3). These amplified CDRs are mixed with oligonucleotides encoding framework regions (step 4), and intact genes encoding variable light and heavy domains are synthesized, using overlap extension PCR (step 5). The synthesized variable light and heavy domains are joined together to form a complete gene encoding scFv antibody fragment with de novo gene sequences (step 6).

transformed into *Escherichia coli* by electroporation. The resulting library contained 75–80% clones with an open reading frame (ORF), as determined by western blot analysis and by sequencing. Restriction enzyme digestion of 32 scFv genes using *Bam*HI and *Bst*NI and sequencing of 52 clones demonstrated that the library exhibited a high degree of variability (data not shown).

In the present investigation we demonstrate the functionality of the molecular design approach of varying all six CDRs and selecting a variety of specific antibody fragments with novel gene sequences, constructed using gene segments from antibodies of different germline origins.

Genetic diversity using in vivo-formed gene segments. The use of in vivo-formed and proofread gene segments has several advantages compared to in vitro-designed and constructed synthetic oligonucleotides employed in the creation of diverse gene libraries (e.g. refs 4, 5, 18, 19). Proofread segments are optimized with regard to functionality and will most likely not encode antigenic T-cell epitopes, since they have been adapted and processed by the immune system.

Analysis of 10 selected scFvs from the n-CoDeR library, specific for different antigens, did not show an increased presence of T-cell epitopes (see Experimental Protocol), as compared to variable domains from 10 normal human antibodies of IgG type, specific for nonhuman proteins (summarized in ref. 14). In contrast, the analysis indicated that scFv selected from n-CoDeR contained fewer T-cell epitopes than the natural antibody variable regions (Table 1). A possible explanation for this is that the natural antibodies contained

Table 1. Analysis of potential T-cell epitopes in scFv-fragments sequences from n-CoDeR using peptide threading[a]

A. Results for heavy chain

n-CoDeR	No. of epitopes	Mature VH	No. of epitopes
Y8	3	GF4H	4
STR5	4	KL4H	7
X4	2	X753	7
M16	3	S55290H	1
L8	7	S55287H	6
L1	0	M979	5
F5	4	L26906H	6
CT173	4	L268	5
SM159	4	L08090H	4
HSP17	4	L08083H	9
Average	3.5	Average	5.4

B. Results for light chain

n-CoDeR	No. of epitopes	Mature VL	No. of epitopes
Y8	4	D84136L	5
STR5	1	M97922L	9
X4	3	D84138L	2
M16	3	S55292L	2
L8	4	75387L	2
L1	4	K14L	1
F5	2	L26899L	1
CT173	3	GF4L	5
SM159	2	L26896L	3
HSP17	4	S55289L	6
Average	3	Average	3.6

[a]Peptide threading analyzes the potential binding of every possible 13-mer from the proteins of interest to human HLA-DR allotypes. The linker was removed before analyses[39].

somatic mutations in their framework regions, whereas, by design, the framework regions of the n-CoDeR-derived antibody fragments were all germline. Importantly, no T-cell epitopes were generated as a function of mixing CDRs from other germlines with the master framework, whereas inclusion of the standard $(G_4\text{-}S)_3$ linker between VH and VL often resulted in the generation of T-cell epitopes (data not shown). It is thus possible that specific binders derived from n-CoDeR and converted into whole antibodies would exhibit even lower immunogenicity compared to antibodies from other phage libraries. In contrast to biologically derived segments, the design of synthetic oligonucleotides is based on our current incomplete knowledge of the variation in the immunoglobulin repertoire, which may result in nonoptimal in vitro variation, including the possibility of encoding antigenic epitopes. Synthetic oligonucleotides are also associated with incomplete synthesis, resulting in deletions and/or mutations, which will reduce the frequency of functional clones. To overcome this problem, the oligonucleotides used to create the n-CoDeR were produced by polymerase chain reaction (PCR), rather than chemically synthesized, which increases the frequency of functional genes (Kobayashi and Söderlind, unpublished data)[11]. The concept of DNA recombination using DNA shuffling technology has been described previously[20]. This technique was employed for the random combination of in vivo-formed gene sequences[21,22]. However, the size and location of DNA to be recombined was undefined. In contrast, the n-CoDeR technology allows sequences of defined size and location to be combined. However, both these examples point to the advantage of using highly functional, in vivo-formed DNA sequence in gene library constructions.

Selection of functional antibodies from n-CoDeR. The designed library was used for the selection of several specific antibody fragments against haptens, peptides, proteins, and carbohydrates (Table 2). Selections were carried out using biotinylated antigen and subsequent capture on streptavidin-coated magnetic beads[6,23]. The specificity of selected clones was confirmed by phage enzyme-linked immunosorbent assay (ELISA) against several unrelated antigens.

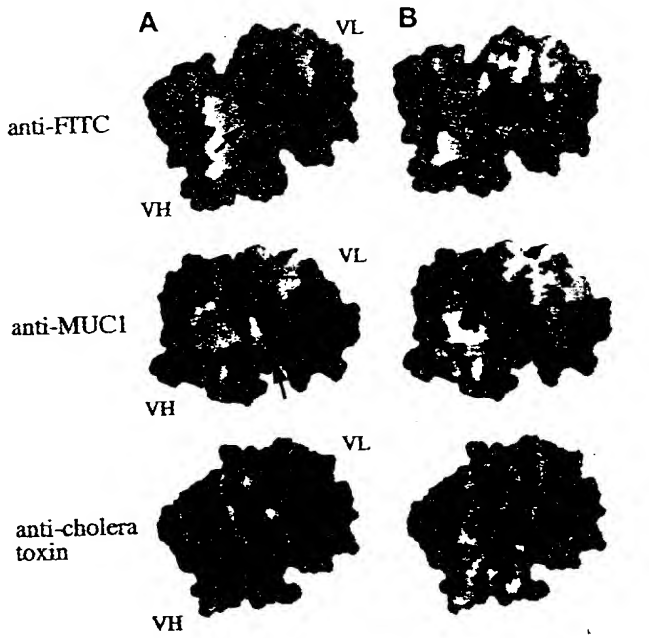

Figure 2. Top view of the scFv antibody fragments specific for fluorescein isothiocyanate (FITC; clone FITC8), MUC1 peptide (Muc159), and cholera toxin (β-subunit) (clone CT17) showing the molecular surface of the individual combining sites. The homology models were generated using the program MODELER[36], and templates were from the Brookhaven Protein Databank (see Experimental Protocol) (A) The surface potential of the scFv fragments was mapped onto the models. Positive potential (≥15 mV) is colored blue, neutral potential (0 mV) is colored gray, and negative potential (≤ -15 mV) is colored red. Arrows mark potential antigen-binding sites. (B) CDR residues, identified as frequent contact residues from x-ray crystallographic analysis of a large set of antigen–antibody complexes[27], were mapped onto the models. The CDRs are color-coded as follows: L1, dark green; L2, blue; L3, magenta; H1, red; H2, light green; H3, cyan. All the figures were generated with GRASP[40].

Table 2. Selection parameters of clones from n-CoDeR library[a]

Antigen	No. of selections	Positive clones/ tested clones	No. of unique clones/ sequenced clones
β-Galactosidase	3	49/59	1/3
Cholera toxin (β-subunit)	3	28/55	3/5
FITC	3	85/87	5/5
Human cell surface antigen	4	10/120	5/10
Human leptin	3	8/60	5/5
Human prostate-specific antigen	4	6/120	4/4
Lewis X	3	4/53	3/3
Lewis Y	3	2/120	2/2
MUC1	3	51/169	6/6
Phenyloxazolone	3	55/59	N/D
Streptavidin	2	7/14	4/4

[a]N/D, not determined.

Table 3. Dissociation constant (K_d) of some monomeric scFv antibodies selected from the n-CoDeR library[a]

Clone scFv	K_d ($\times 10^{-9}$ M^{-1})
FITC8	0.9
FITC19	1.9
FITC18	2.0
FITC9	3.2
CT17	3.1
CT15	7.6
MUC41	67
MUC23	123
MUC46	200
MUC20	270
MUC44	420

[a]CT, cholera toxin (β-subunit); FITC, fluorescein isothiocyanate; MUC, mucin-1.

Most of the antigens gave high frequencies of binders after three rounds of selection. Binders against carbohydrate antigens could also be selected, albeit at a somewhat lower frequency. Affinities for a set of selected antibody fragments were then determined using the monomeric form of the scFv antibody fragment (Table 3).

A central question in the evaluation of n-CoDeR was whether antibody fragments encoded by the novel genes would be functional with respect to antigen binding. Our library design was based on the principle that a specific single framework had the potential to include CDRs derived from other germline gene sequences. However, structural incompatibility between these foreign CDRs and the fixed framework could potentially prevent the formation of functional antibody-combining sites. This was not found to be the case. A number of different antibody specificities could be generated, demonstrating functional structural diversity in the antigen-combining sites. An analysis using homology studies (see Experimental Protocol) revealed that the three-dimensional structure of three specific n-CoDeR antibodies exhibited an overall immunoglobulin-like fold. This fold was verified to be structurally most similar to antibody molecules according to the Dali program[24]; the best matches giving a Z-score of ~16.0 and an r.m.s. deviation for Cα rigid-body superimposition of ~0.6 Å. Analysis of the solvent-accessible surface showed that the surface potential and in particular the topology of the combining sites differed among the three scFv fragments (Fig. 2A). The relationship between the topography of the combining site and the type of antigen recognized has emerged from previous x-ray crystallographic studies of antigen–antibody complexes[25]. At least three classes of sites binding to haptens (cavity), peptides (groove), and proteins (planar) have been described. When analyzing the accessible surface of our three models in more detail, we found a deep cavity within the center of the anticipated combining site of the anti-hapten antibody, fluorescein isothiocyanate (FITC) and an open groove on the anti-peptide antibody (MUC1 peptide), whereas no obvious site could be identified on the anti-protein antibody against cholera toxin (β-subunit) (Fig. 2A). Several positively charged (Fig. 2A) and aromatic residues (data not shown), known from x-ray crystallographic studies of an anti-FITC antibody to be of importance for binding of antigen[26], are present within the deep cavity detected on the anti-FITC antibody. To further define the potential antigen-binding site, we mapped putative contact residues within the CDRs[27] onto the modeled structure of these three scFv fragments (Fig. 2B). A potential combining site is clearly highlighted within the deep cavity observed on the anti-FITC scFv, and the cavity is almost completely formed with these residues. In the case of the anti-MUC1 scFv, the CDR residues cluster more along an open groove and thus may constitute a direct contact site for the peptide. In contrast, a considerably larger and apparently flatter, more undulating surface area is highlighted on the anti-cholera toxin scFv, which is indicative of the much larger part of the paratope generally used for binding to a protein antigen instead of a hapten (600–900 Å^2 vs. 100–300 Å^2).

The view of high functionality of n-CoDeR is supported by the fact that subnanomolar affinities can be directly selected from our library, a performance previously only reported from larger libraries[28–30]. The fact that carbohydrate clones specific for Lewis Y and Lewis X, previously found only in an immunized antibody gene library[31], can be directly selected from the library, points to large functional genetic variation. Finally, the selected VH framework is

Table 4A. Germline gene origin of CDRs found in unselected clones of the n-CoDeR library[a]

Clone	CDRH1		CDRH2		CDRH3	CDRL1		CDRL2		CDRL3	
	Origin	No. of nt mut	Origin	No. of nt mut	No. of aa	Origin	No. of nt mut	Origin	No. of nt mut	Origin	No. of nt mut
O3	3-7	0	3-30, 30-30-5	0	16	1-44	3	1-40, 1-50	0	1-36, 1-44	0
O9	3-9, 3-43	0	3-11, 3-48	8	14	1-44	1	1-44, 1-47	4	1-36, 1-44	0
O11	3-30, 3-30-3, 3-64	0	3-7	1	8	1-44	0	1-51	5	1-40	0
O12	3-30, 3-30-5, 3-33	0	3-23	0	12	1-40	0	1-44	2	1-47	3
O20	3-30, 3-30-3, 3-47, 3-64	2	3-23	5	10	1-36	6	1-44	3	1-36, 1-44	0
022	3-48	1	3-7	8	10	1-47	5	1-44	2	1-47	0
O23	3-11, 3-13	4	3-11, 3-23, 3-48	14	20	1-40	5	1-47	1	1-47	3
O25	3-74	2	3-23	9	13	1-44	2	1-40, 1-50	3	1-40	0
O29	3-30, 3-30-5, 3-33	1	3-23	10	9	1-47	6	1-44	1	1-47	0
O32	3-23	1	3-21	0	7	1-44	5	1-44	0	2-11	2

[a]Pseudogenes with sequence similarity identical to functional or ORF variants have not been included if such genes exist. Orphan VH loci found on chromosomes 15 and 16 were not considered in this comparison. The length of the amplified sequences are as follows: CDRH1, 19 nt; CDRH2, 78 nt; CDRL1, 41 nt; CDRL2, 21 nt. nt, Nucleotides; mut, mutations; aa, amino acids.

Table 4B. Germline gene origin of CDRs found in selected clones of the n-CoDeR library[a]

Clone	Specificity	CDRH1		CDRH2		CDRH3	CDRL1		CDRL2		CDRL3	
		Origin	No. of nt mut	Origin	No. of nt mut	No. of aa	Origin	No. of nt mut	Origin	No. of nt mut	Origin	No. of nt mut
PSA22	Prostate-specific antigen	3-48	2	3-30, 3-30-5	4	12	1-44	6	1-47	0	1-47	4
CT9	Cholera toxin β-subunit	3-23	4	3-7	11	14	1-44	2	1-40, 1-50	3	1-36, 1-44	2
CT17	Cholera toxin β-subunit	3-60, 3-62[b]	0	3-23	1	20	1-36, 1-44	6	1-40, 1-50	0	1-36, 1-44	0
FITC8	Fluorescein	3-7	1	3-23	8	8	1-40	0	1-40, 1-50	1	1-47	0
FITC19	Fluorescein	3-11	0	3-23	8	8	1-36	2	1-47	0	1-47	1
Lep1	Leptin	3-49	1	3-7	1	11	1-44	3	1-36	0	1-36, 1-44	6
Lep4	Leptin	3-20	0	3-23	1	14	1-36, 1-44	11	1-40, 1-50	0	1-47	0
MUC20	MUC1 tandem repeat	3-11	1	3-30, 3-30-5	5	9	1-36, 1-44	6	1-40, 1-50	1	1-36, 1-44	0
MUC46	MUC1 tandem repeat	3-7	0	3-30-3	5	9	1-44	6	1-51	1	1-47	4
Strep6	Streptavidin	3-11	1	3-23	2	13	1-44	0	1-41, 1-51	0	2-18	1

[a]Pseudogenes with sequence similarity identical to functional or ORF variants have not been included if such genes exist. Orphan VH loci found on chromosomes 15 and 16 were not considered in this comparison. The lengths of the amplified sequences are as follows: CDRH1, 19 nt; CDRH2, 78 nt; CDRL1, 41 nt; CDRL2, 21 nt. nt, Nucleotides; mut, mutations; aa, amino acids.
[b]Both of these germline genes are pseudogenes.

DP47, which belongs to the VH3 family, the members of which have been shown to have limited polyreactivity compared with members of other VH families[32]. This fact might also contribute to the functionality and specificity of the selected antibody fragments.

CDR-usage in the antibody fragments. We sequenced a total of 53 clones from the library that had not been selected for antigen binding, and found all of them to be mosaic; i.e. each clone contained combinations of CDR sequences of different germline origin (Table 4A). Some of the CDR sequences in the library were found to originate from unmutated germline sequences, whereas some sequences contained mutations. The frequencies and locations of these mutations exclude the PCR process as the predominant source of nucleotide substitutions (data not shown). Thus, the functional genetic variation in n-CoDeR arises from both the combination of CDRs originating from different germline sequences and by the presence of mutations in individual CDR sequences. These mutations are a consequence of the somatic mutation process. Furthermore, after sequencing the clones selected for antigen specificity, the antibody fragment genes were all of the expected size, without stop codons or deletions, and were also found to be mosaic (Table 4B). In no clone was the full-length original DP47 and DPL3 germline sequence combination found. From the analyses of the selected clones, it is clear that the immunoglobulin scaffold has a thus far uncharacterized ability to harness foreign CDR loops. This incorporation of foreign CDRs into a single framework has similarities to the gene conversion event in some species, which allows for translocation of defined gene segments into a given gene sequence[33]. No such in vivo mechanism has been described in the human immune system, and the flexibility of the immunglobulin scaffold to accept foreign gene sequences presented here demonstrates that functional human antibodies may be formed utilizing principles other than what evolution provides.

Experimental protocol

Isolation of in vivo proofread CDRs from various B-cell sources. To achieve largest possible variation in the library, B cells originating from different individuals and from different lymphoid tissue/organs, were used as a source of CDRs. RNA was prepared from lymph nodes from seven different donors, peripheral blood mononuclear cells (PBMCs) from nine different donors, spleen from one donor, and tonsil RNA from five donors. The total number of B cells used was approximately 2×10^9. Total RNA was prepared using an RNA isolation kit from Promega (Madison, WI), and cDNA synthesis was performed with Superscript II (Life Technologies Inc., Paisley, UK) according to the method of Engberg and coworkers[17], using 50 μg RNA per reaction. The cDNA synthesis reaction was primed with 3′ primers specific for the human γ, μ, α, and λ constant region. A total of 30 reactions (15 VH and 15 VL) were performed, and to avoid bias all samples were treated separately in the following PCR amplification steps. Primers used in the PCR amplification of the heavy-chain CDRs were essentially as described[11], and the primers used for the amplification of light-chain CDRs were designed according to the same principle. For each CDR a pair of primers was used, homologous to the master framework. These primers allowed for the amplification of CDRs from a variety of germline sequences. The CDRs are defined as described in VBASE (http://www.mrc-cpe.cam.ac.uk/imt-doc/public/INTRO.html). Each of the six CDRs was PCR amplified using TaqPlus Precision PCR System (Stratagene, La Jolla, CA), according to the manufacturer's instructions.

Library construction. PCR amplification of framework regions was performed using the same principle as for the CDRs, and a clone containing the DP47 and DPL3 frameworks[12] was used as template. Purification of the PCR product from both CDR and framework amplifications were performed essentially as described[11]. The single-stranded CDR and framework fragments were assembled into scFv genes in a two-step overlap extension PCR, using TaqPlus Precision PCR System, as described[12]. The assembled scFv antibody gene fragments were purified using QIAquick PCR Purification Kit (Qiagen, Hilden, Germany), digested with *Sfi*I and *Not*I (New England Biolabs, Beverly, MA), and purified using QIAquick Gel Extraction Kit (Qiagen). They were then ligated into the *Not*I/*Sfi*I sites in the pFAB5c-His vector and transformed into electrocompetent TOP10F' *E. coli* cells. Amplification of the library and phage stock production with R408 helperphage was done as described[17].

Selection of specific antibody fragments. Selection on biotinylated antigen was done using streptavidin-coated Dynabeads (Dynal A.S., Oslo, Norway), as described[6,23], except that bound phage were eluted with trypsin, as described[17]. The amount of phage used was 1.5×10^{13} c.f.u. in the first selection and approximately 1×10^{10} in later selection steps. Antigen concentrations ranged from 1×10^{-7} M to 1×10^{-12} M. For phagemid amplification, log phase *E. coli* TOP10F' cells were infected with the eluted phage and grown in Luria–Bertani (LB) medium containing ampicillin (100 μg ml^{-1}) and tetracycline (15 μg ml^{-1}) at 37°C for ~3 h before infection with 20 times excess of helper phage R408. Isopropyl-β-D-thiogalactoside (IPTG) was added to 0.1 mM, and growth was continued at 25°C for 13 h before the amplified phage were harvested. After selections, phage displaying antigen-binding antibody fragments were identified by ELISA using an anti-M13 peroxidase-conjugated antibody (Pharmacia Biotech, Uppsala, Sweden), as described[34].

Analyses of CDR origin. DNA-sequencing reactions were performed using BigDye Terminator Sequencing kit (Applied Biosystems, Warrington, UK), according to the manufacturer's protocol, and were gel analyzed by CyberGene AB (Stockholm, Sweden). Analysis of the CDR origin was performed using the FastA algorithm run on the Wisconsin Package (version 8: Genetics Computer Group, Madison, WI). Comparisons of sequences amplified by the primers used in these experiments were made with sequences found in the VBASE sequence directory (http://www.mrc-cpe.cam.ac.uk/imt-doc/public/INTRO.html). Each CDR was assigned a locus origin as defined by the description of the human Ig VH and Vλ loci[35,36].

Affinity determination. The monomeric form of the antibody fragments was purified by size-exclusion gel chromatography on a Superose G75 FPLC 16/60 column (Amersham Pharmacia Biotech AB, Sweden). The affinity was determined as described[37].

Structure modeling. A homology model was generated for the scFv antibody fragments specific for FITC, MUC1 peptide, and cholera toxin β-subunit) (CT) using the program MODELER[38]. Templates were selected from the Brookhaven Protein Databank: 1IgM (FITC, CT) and 1AQK (MUC1) for the VH domain and 2FB4 (FITC, MUC1, CT) for the VL domain. Tags were removed before model generation. The models were built by Molecular Simulations Inc. (San Diego, CA).

Analysis of T-cell epitopes. Peptide threading[39] was used to analyze the presence of potential T-cell epitopes in VH and VL domains of 10 selected scFv fragments. The analyses were performed at Biovation Ltd. (Aberdeen, Scotland). For comparison, the variable domains from 10 natural human antibody sequences were analyzed. Peptide threading analyzes the potential binding of every possible 13-mer from the proteins of interest to human HLA-DR allotypes. The following HLA-DR allotypes, which represent >96% of the population, were included in the analysis; 1, 15(2), 16(2), 17(3), 18(3), 4, 11(5), 12(5), 13(6), 14(6), 7, 8, 9, 10, 52, 53, 103.

Acknowledgments

This investigation was supported by grants from The National Research Council for Engineering Sciences, the National Board for Laboratory Animals, and in part by the BIOSAM program of NUTEK and the European Commission DGXII.

1. Barbas, C.F., Kang, A.S., Lerner, R.A. & Benkovic, S.J. Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. *Proc. Natl. Acad. Sci. USA* **88**, 7978–7982 (1991).
2. Clackson, T., Hoogenboom, H.R., Griffiths, A.D. & Winter, G. Making antibody fragments using phage display libraries. *Nature* **352**, 624–628 (1991).
3. Marks, J.D. et al. By-passing immunization. Human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.* **222**, 581–597 (1991).
4. Barbas, C.F., Bain, J.D., Hoekstra, D.M. & Lerner, R.A. Semisynthetic combinatorial antibody libraries: a chemical solution to the diversity problem. *Proc. Natl. Acad. Sci. USA* **89**, 4457–4461 (1992).
5. Hoogenboom, H.R. & Winter, G. By-passing immunisation. Human antibodies from synthetic repertoires of germline VH gene segments rearranged in vitro. *J. Mol. Biol.* **227**, 381–388 (1992).
6. Griffiths, A.D. et al. Isolation of high affinity human antibodies directly from large synthetic repertoires. *EMBO J.* **13**, 3245–3260 (1994).
7. Hayashi, N. et al. Simultaneous mutagenesis of antibody CDR regions by overlap extension and PCR. *Biotechniques* **310**, 312–315 (1994).
8. de Kruif, J., Boel, E. & Logtenberg, T. Selection and application of human single chain Fv antibody fragments from a semi-synthetic phage antibody display library with designed CDR3 regions. *J. Mol. Biol.* **248**, 97–105 (1995).
9. Söderlind, E., Vergeles, M. & Borrebaeck, C.A.K. Domain libraries: synthetic diversity for *de novo* design of antibody V-regions. *Gene* **160**, 269–272 (1995).
10. Pini, A. et al.Design and use of a phage display library. Human antibodies with subnanomolar affinity against a marker of angiogenesis eluted from a two-dimensional gel. *J. Biol. Chem.* **273**, 21769–21776 (1998).
11. Jirholt, P., Ohlin, M., Borrebaeck, C.A.K. & Söderlind, E. Exploiting sequence space: shuffling in vivo formed complementarity determining regions into a master framework. *Gene* **215**, 471–476 (1998).
12. Kobayashi, N., Söderlind, E. & Borrebaeck, C.A.K. Analysis of assembly of synthetic antibody fragments: expression of functional scFv with predefined specificity. *Biotechniques* **23**, 500–503 (1997).
13. Huang, S.C., Jiang, R., Glas, A.M. & Milner, E.C Non-stochastic utilization of Ig V region genes in unselected human peripheral B cells. *Mol. Immunol.* **33**, 553–560 (1996).
14. Ohlin, M. & Borrebaeck, C.A.K. Characteristics of human antibody repertoires following active immune responses in vivo. Mol. Immunol. **33**, 583–592 (1996).
15. Ignatovich, O., Tomlinson, I.M., Jones, P.T. & Winter, G. The creation of diversity in the human immunoglobulin V(lambda) repertoire. *J. Mol. Biol.* **268**, 69–77 (1997).
16. de Wildt, R.M., Hoet, R.M.A., van Venrooij, W.J., Tomlinson, I.M. & Winter, G. Analysis of heavy and light chain pairings indicates that receptor editing shapes the human antibody repertoire. J. Mol. Biol. **285**, 895–901 (1999).
17. Engberg, J. et al. Phage-display libraries of murine and human antibody Fab fragments. *Mol. Biotechnol.* **6**, 287–310 (1996).
18. Crameri, A. & Stemmer, W.P. Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wild-type sequences. *Biotechniques* **18**,194–196 (1996).
19. Virnekäs, B. et al.Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Res.* **22**, 5600–5607 (1994).
20. Stemmer, W.P. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389–391 (1994).
21. Crameri, A., Raillard, S.A., Bermudez, E. & Stemmer, W.P. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288–291 (1998).
22. Chang, C.C. et al. Evolution of a cytokine using DNA family shuffling. *Nat. Biotechnol.* **8**, 793–797 (1999).
23. Hawkins, R.E., Russell, S.J. & Winter, G. Selection of phage antibodies by binding affinity. Mimicking affinity maturation. *J. Mol. Biol.* **226**, 889–896 (1992).
24. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
25. Webster, D.M., Henry, A.H. & Rees, A.R. Antibody–antigen interactions. *Curr. Opin. Struct. Biol.* **4**, 123–129 (1994).
26. Herron, J.N., He, X.M., Mason, M.L., Voss, E.W. Jr. & Edmundson, A.B. Three-dimensional structure of a fluorescein–Fab complex crystallized in 2-methyl-2,4-pentanediol. *Proteins* **5**, 271–280 (1989).
27. MacCallum, R.M., Martin, A.C. & Thornton, J.M. Antibody–antigen interactions: contact analysis and binding site topography. *J. Mol. Biol.* **262**, 732–745 (1996).
28. Vaughan, T.J. et al. Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nat. Biotechnol.* **14**, 309–314 (1996).
29. Sheets, M.D. et al. Efficient construction of a large nonimmune phage antibody library: the production of high-affinity human single-chain antibodies to protein antigens. *Proc. Natl. Acad. Sci. USA* **26**, 6157–6162 (1998).
30. de Haard, H.J. et al. A large non-immunized human Fab fragment phage library that permits rapid isolation and kinetic analysis of high affinity antibodies. *J. Biol. Chem.* **274**, 18218–18230 (1999).
31. Mao, S. et al. Phage display library selection of high-affinity human single-chain antibodies to tumor-associated carbohydrate antigens sialyl Lewisx and Lewisx. Proc. Natl. Acad. Sci. USA **96**, 6953–6958 (1999).
32. Ditzel, H.J., Itoh, K. & Burton, D.R. Determinants of polyreactivity in a large panel of recombinant human antibodies from HIV-1 infection. *J. Immunol.* **157**, 739–749 (1996).
33. Meyer, A., Parng, C.L., Hansal, S.A., Osborne, B.A. & Goldsby, R.A. Immunoglobulin gene diversification in cattle. Int. Rev. Immunol. **15**, 165–183 (1997).
34. Söderlind, E. et al. Chaperonin-assisted phage display of antibody fragments on filamentous bacteriophages. *Biotechnology* **11**, 503–507 (1993).
35. Pallarès, N., Frippiat, J.P., Giudicelli, V. & Lefranc, M.-P. The human immunoglobulin lambda variable (IGLV) genes and joining (IGLJ) segments. *Exp. Clin. Immunogenet.* **15**, 8–18 (1998).
36. Pallarès, N., Lefebvre, S., Contet, V., Matsuda, F. & Lefranc M.-P. The human immunoglobulin heavy variable genes. *Exp. Clin. Immunogenet.* **16**, 36–60 (1999).
37. Friguet, B., Chaffotte, A.F., Djavadi-Ohaniance, L. & Goldberg, M.E. Measurements of the true affinity constant in solution of antigen–antibody complexes by enzyme-linked immunosorbent assay. *J. Immunol. Methods* **77**, 305–319 (1985).
38. Sali, A. & Blundell, T.L. Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
39. J. Fothergill, Kemp, G., Brooks, T. & Carr, F. Identification of MHC binding peptides Patent application number WO 98/59244.
40. Honig, B. & Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **268**, 1144–1149 (1995).